GIS workflow for continuous soil moisture estimation based on medium resolution satellite data

Boudewijn van Leeuwen
Dep. of Physical Geography and Geoinformatics
University of Szeged
Egyetem utca 2-6, 6722
Szeged, Hungary
leeuwen@geo.u-szeged.hu

Abstract

Climate models predict a combined trend of higher average temperatures and less summer precipitation for the Carpathian Basin. This makes the region vulnerable to future droughts. Decreasing soil moisture is an important indicator for drought and therefore it is important to develop a method that allows for its continuous monitoring at regional scale. This study presents the development of an automatic workflow for satellite based soil moisture estimates which are validated using in situ ground measurements. Pre-processed MODIS normalized vegetation index maps are reclassified into maps with 10 normalized vegetation classes. For these areas, the land surface temperature is calculated based on the thermal data from the same MODIS instrument acquired at the same time. This way, for every vegetation class, temperature statistics are calculated and a linear relationship between the land surface temperature and soil moisture is determined based on the assumption that for equal vegetation classes the soil surface temperature is primarily dependent on the soil moisture content. This results in daily soil moisture index (SMI) maps.

Regression analysis is carried out to calibrate the relative SMI values with *in situ* soil moisture measurements at measurement stations and to derive soil moisture values. Continuous calculation of soil moisture provides trend information, which can help to predict future periods of drought.

Keywords: soil moisture, land surface temperature, NDVI, MODIS.

1 Introduction

Soil moisture is a parameter that via its role in the global energy and water cycle is important in many natural and agricultural processes [5, 6, 8]. Its measurement is performed as scattered point measurements at a discrete interval. Many applications would benefit from detailed, spatially and temporarily continuous data on soil moisture over large areas over long periods. With *in situ* measurements, this is not feasible due to financial and physical constraints. Soil moisture derived from satellite data provides the opportunity to overcome these constraints. Many methods to calculate soil moisture have been developed and applied, but a single best algorithm has not been determined yet [1]. Two main methodologies in spaceborne remote sensing based soil moisture observation can be identified. The first one is based on measurements in the microwave part of the electromagnetic spectrum, while the second is based on thermal, visible and infrared observations[8].

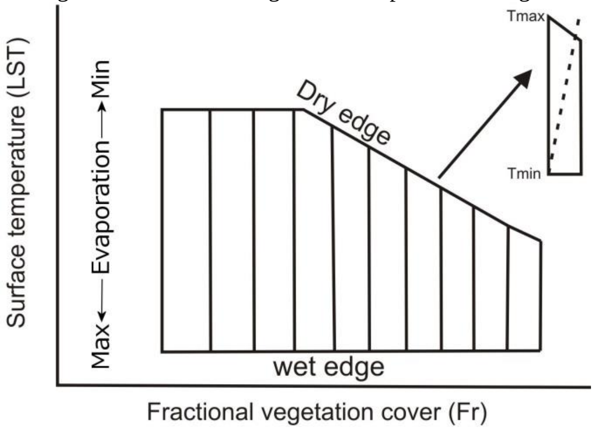
The microwave based method uses the large difference in the dielectric properties of liquid water (~80) and dry soil (<4). This difference of the soil's dielectrical constant results in a variation in emissivity from 0.95 for dry soils to 0.6 or less for wet soils, with changes of the corresponding magnitude for the soil's reflectivity [7]. Two types of microwave instruments are applied in soil moisture measurements. The first are passive radiometers that measure the changes in emissivity. The second are active radiometers. These radars emit a pulse and measure the back scattered return which is a function of the soil's reflectivity.

Microwaves have a wavelength between 1 mm and 1 meter. Due to their longer wavelengths, they have the capability to penetrate clouds, haze and rain which allows the method to be applied in many circumstances. However, the applicability of measuring soil moisture by microwave sensors is limited due to the poor spatial resolution of passive microwave and the lower temporal resolution and strong sensitivity of active microwave instruments to vegetation cover and surface roughness [6].

The second method to determine soil moisture is based on a combination of vegetation data derived from visible and near infra-red satellite data and thermal data. For every pixel the vegetation cover is derived and the surface temperature is calculated. Plotting these values in a two dimensional vegetation - temperature space results in a theoretical figure that resembles a triangle (Figure 1). On non-vegetated soils and in full vegetation areas, evaporation and transpiration increase as the water content rises. When soils are moist, the latent heat fluxes increase because of the greater absorption of water. This process causes sensible heat to decrease. In dry soils the process is the inverse of this. The radiative energy is not consumed in the evapotranspiration process, and the sensible heat increases, raising the surface temperature [8]. The triangle is only fully developed if all vegetation classes are available in the study area. The method is valid when both minimum and maximum surface soil wetness can be observed within the geographical extent of the study area. This assumption requires a heterogeneous study area with uniform atmospheric conditions. The major disadvantages of the triangle method are its dependency on cloud free data and required distribution of vegetation classes [5].

For this study, the triangle method was selected because of the readily available base data and the lack of auxiliary data required for the calculations [4]. This makes it also relatively easy to automate the workflow.

Figure 1: Theoretical vegetation-temperature triangle.



2 Data

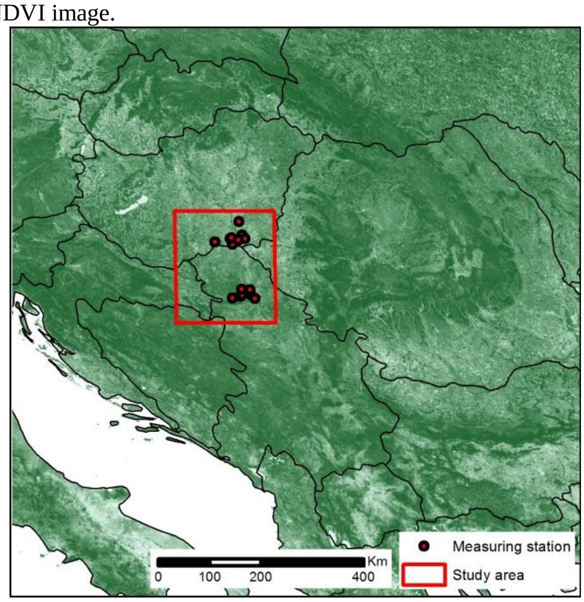
In this study the Moderate Resolution Imaging Spectroradiometer (MODIS) MOD11A1 and MOD13Q1 data products are used to calculate soil moisture and soil moisture index maps. Both products are derived from the same base data, and are therefore collected at the same time, from the same area and with the same geometry. The data can be downloaded automatically and free of charge from the USGS Earth Explorer website.

The first product is the Land Surface Temperature and Emissivity (MOD11A1) product, which is calculated on a daily basis. Among others, it contains a land surface temperature (LST) layer with a spatial resolution of 1000 meter and data quality layer [9]. The second product is the Vegetation Indices (MOD13Q1) data set, which contains a Normalized Difference Vegetation Index (NDVI), an Enhanced Vegetation Index and a data quality layer. The NDVI layer is a composite product storing the maximum NDVI value within a 16 days interval for each pixel with a spatial resolution of 250 meter [3]. The 16 day interval provides enough data to almost continuously generate NDVI values for every pixel, while the interval is short enough to represent the changes in NDVI without smoothing them out too much. Both data product have been radiometrically and atmospherically corrected by USGS [3, 9].

The study area is a subset of the original MODIS LST and NDVI images. It extends from the south of Hungary to the Vojvodina region in Serbia, and covers an area of about 200 x 215 kilometres (Figure 2).

The region is mainly agricultural although there are several large cities well. The climate is moderate continental, with cold winters and hot and humid summers with a large range of extreme temperatures and non-equal distribution of rainfall per months making the area susceptible to floods as well as drought [2]. A large variety of soils can be found in the region, ranging from blown sand and alluvial meadow soils in the north to chernozem in the south. The Danube and Tisza rivers are the main waterways in the area. Several large lakes serve as ecological buffers and sources for recreation and agricultural needs.

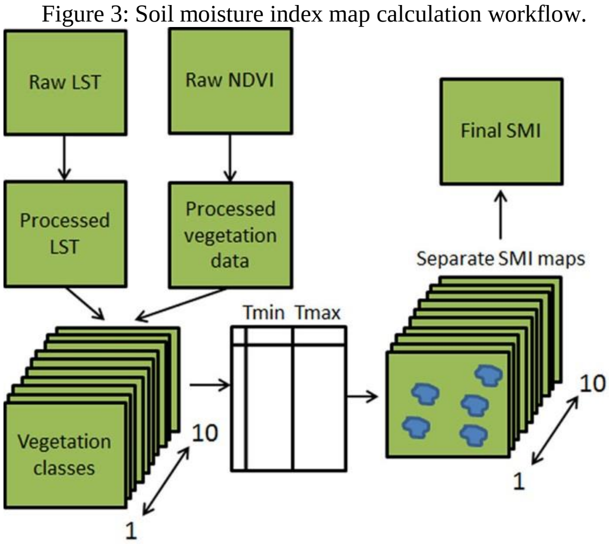
Figure 2: Study area and the area covered by the MODIS NDVI image.



3 Methodology

The workflow to produce SMI maps consists of four major steps: (1) downloading of the data sets, (2) processing of the NDVI data, (3) processing of the land surface temperature data, (4) creation of the LST-NDVI triangle and (5) Calculation of the soil moisture index (SMI) maps. These steps have been implemented as Python scripts using ArcGIS geoprocessing tools (Figure 3).

Figure 3: Soil moisture index map calculation workflow.



1. In the first step, the MOD11A1 and the MOD13Q1 data is automatically downloaded from the USGS Earth Explorer database by a Python script that uses the ArcGIS geoprocessing library. Based on a start date, end date, a shape file defining the spatial extent and some definitions of output names, the script automatically

selects and downloads the required LST data and the corresponding NDVI image.

2. The processing of the NDVI data starts with the creation of a mask based on the quality layer in the MOD13 dataset. Only NDVI data of sufficient quality is extracted from the original dataset. From this data, a spatial subset is created based on the study area. The subset is calibrated to return to the original -1 to 1 data range for NDVI values. These values are then normalized to get the so called vegetation fraction *F* using:

$$F = \left(\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}}\right)^2 \quad (1)$$

   where NDVI is the processed NDVI value, $NDVI_{min}$ is the minimum NDVI value, and $NDVI_{max}$ is the maximum value in the image. In regions where all vegetation-cover types are available (from bare soils to full-vegetated), the maximum *F* is associated with 100% vegetation cover.

3. The land surface temperature data is also masked based on the quality layer in the dataset. The extracted data is then spatially subsetted and calibrated. In the final step, the LST data with a spatial resolution of 1000 meter is resampled to the 250 meter resolution of the vegetation fraction *F* data.

4. The vegetation fraction dataset is reclassified to 10 classes with equal class width. For each vegetation class, the minimum and maximum temperature values are derived from the LST data within the area covered by that class. Within each class, the linear relationship is determined between the LST and soil moisture index (SMI) value, using:

$$SMI = \left(\frac{LST_{min} - LST}{LST_{max} - LST_{min}}\right) + 1 \qquad (2)$$

   where LST is the MODIS based land surface temperature, LST_{min} is the minimum temperature in the particular vegetation fraction class, and LST_{max} is the maximum temperature.

5. Based on the established relationships, for every pixel within each of the 10 vegetation fraction classes, a soil moisture index value is calculated. Combining all 10 SMI maps results in an SMI map for the total area. (Figure 6).

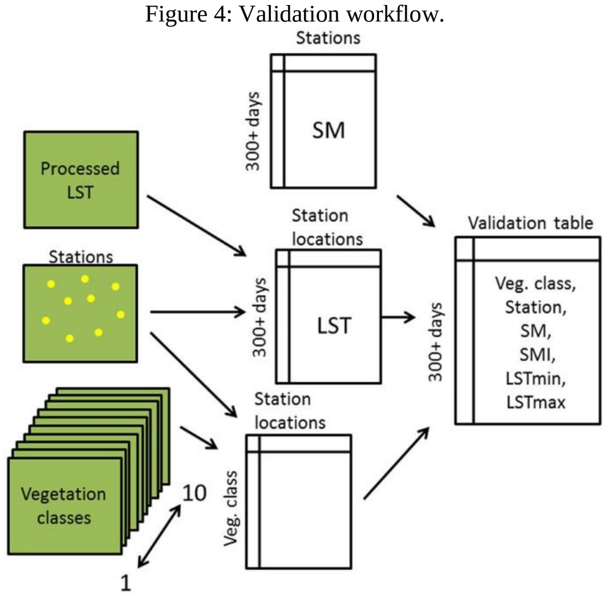
The soil moisture index maps provide a value between 0 and 1 indicating the relative amount of soil moisture within the area, where 0 indicates the lowest soil moisture and 1 means the highest soil moisture on a particular day. Before calibration of the SMI values, it is not possible to perform a quantitative comparison between different days.

3.1 Validation

To validate the calculated SMI values, the linear relationship between the land surface temperature and the soil moisture within one vegetation class is used. At the same time of the satellite land surface temperature measurements, soil moisture measurements were taken using a network of *in situ* measurements stations. If on a particular day, data from multiple stations is available from within one vegetation class, a system of linear equations can be derived and solved. Solving the equations provides the relationship between the measured LST values and the calculated soil moisture values SM_c. When two measurement stations are available in one class, only the linear relationship between the two parameters can be determined. When more measurement stations fall into one vegetation class, the system of equations is overdetermined and the linear relationship as well as the coefficient of determination can be calculated using linear regression. This way it is possible to evaluate the relationships quantitatively. Figure 4 shows the validation workflow, where for a particular day, in each class, the SM value measured at a measurement station and the LST values derived from the satellite data are selected and stored in a validation table. Within this table the linear relationship between the LST and MS values is determined and stored.

When the linear relationship is determined on a particular day, within a vegetation class it is possible to calculate the absolute soil moisture value SM (in v/v) instead of the soil moisture index SMI.

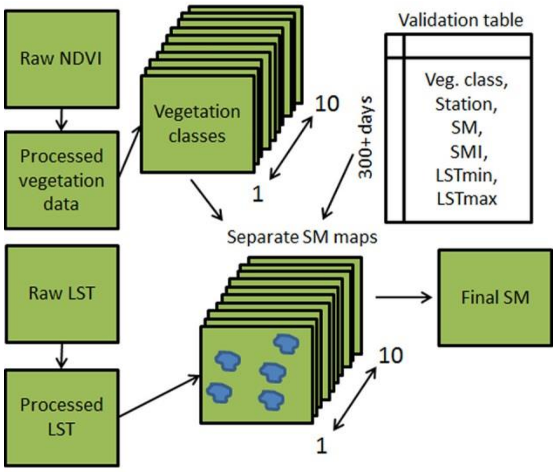
Figure 4: Validation workflow.



3.2 Calibration

The linear relationship determined during the validation is used to calculate the soil moisture maps based on the MODIS LST and NDVI data (Figure 5). The LST and NDVI data are processed similarly as during the soil moisture index maps creation; taking the quality data, spatial extent of the study area and the calibration information into account. For each vegetation class, the particular parameters determining the SM – LST relationship are selected from the table and for each pixel the LST value is converted to a SM value. This results in an SM map for every vegetation class. During the final step, the separate SM maps are combined forming a SM map for the complete area.

Figure 5: Calibration workflow.

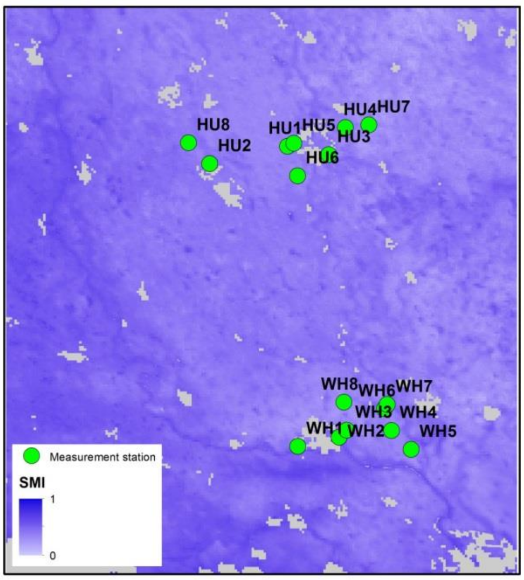


4 Results

Between 1 January 2014 and 4 November 2014, 206 SMI maps could be created. On 102 days, it was not possible to derive the SMI maps because of lack of high quality data. Usually this was due to frost, snow cover or cloud cover preventing the acquisition of land surface temperature images, or the NDVI data was of insufficient quality. In some cases not all vegetation classes were determined, resulting in incomplete definition of the vegetation-temperature triangle. This also prevents the calculation of the SMI map for that day.

The SMI maps offer a qualitative overview of the distribution of soil moisture in the study area (Figure 6.). Obvious are the main rivers with surrounding forests on their floodplains. Also other forested areas in the south (Fruska Gora national park) and west (Gemenc) can clearly be seen.

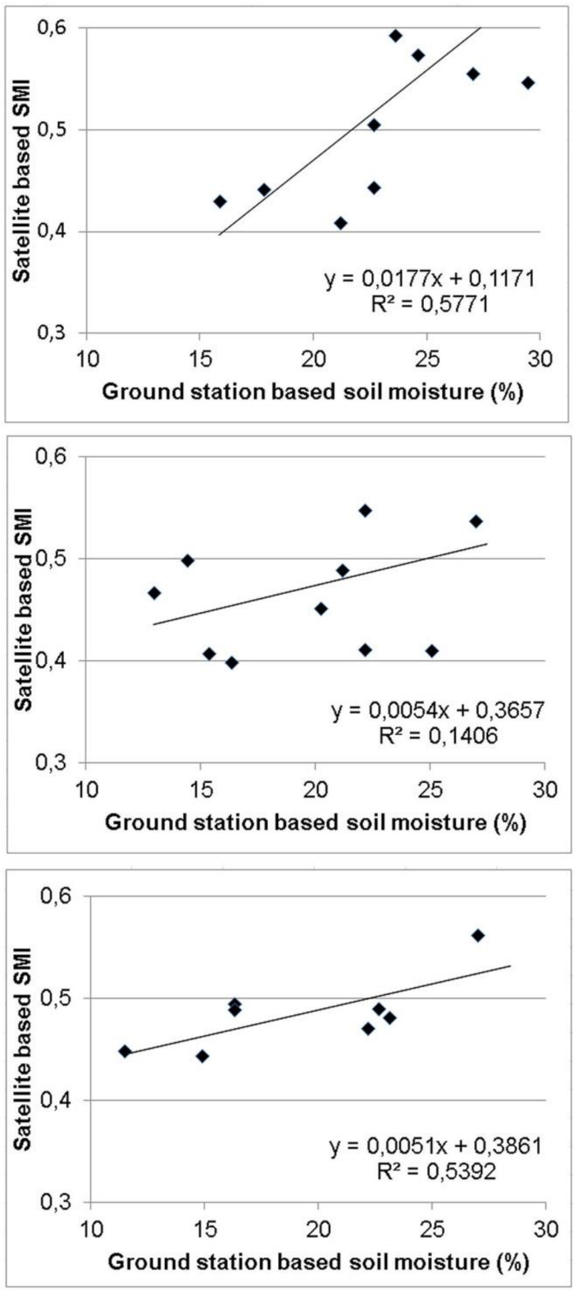
Figure 6: Resulting soil moisture index map of 23 may 2014.



Validation is possible on 115 days, since on those days in at least one vegetation class, data from at least two ground measurement stations is available. In 56 situations, more than 2 *in situ* soil moisture measurements – surface temperature measurements combinations are available, therefore allowing to perform linear regression. The remaining 82 combinations are only used to determine the linear relationship.

The coefficient of determination r^2 shows a large variation, ranging from under 0.1 to above 0.95. Figure 7 shows the relationship between the satellite based SMI and the *in situ* soil moisture values for three different days.

Figure 7: Example of validation results for 3 particular days in 2014.



The average r^2 for all days in the data set and all vegetation classes was 0.40. Several reasons may exist for this low value. First, there is a large difference between the resolution of the satellite measurements and the point measurements on the ground. One pixel of the satellite data may include a variety of vegetation, topography, soil types and other factors influencing the soil moisture values. These may not be the same at the position of the ground measurements station within the pixel. As a consequence, the ground measurements may not be representative for the larger area. Second, the *in situ* measurements are taken at a depth of 10 centimeter, while the satellite derived values are the result of LST and VI data on the surface. The influence of the soil moisture at 10 centimeter on the surface may vary depending on the type and amount of vegetation. Also the influence of precipitation may show up later at the *in situ* measurements than at the satellite measurements depending on the infiltration rate. Finally, the accuracy of the LST values is about 1º K [9]. A small error in the determination of the linear relationship can result in a large coefficient of determination if there are only limited soil moisture measurements – surface temperature measurements combinations available.

Evaluating the coefficient of determination per class does not show a trend between the amount of vegetation and the correlation. Combinations showing an r^2 lower than 0.5 were not used for calibration.

The calibration results are determined for those classes that are defined during the validation, other part of the maps are shown in gray. Figure 8 shows soil moisture maps from two days apart.

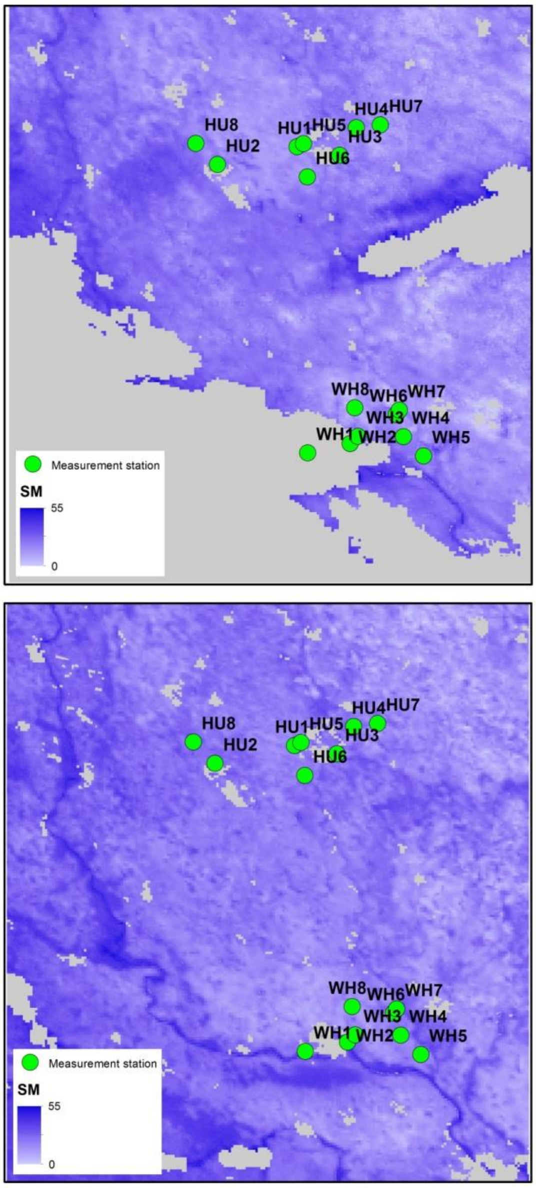
5 Discussion and conclusions

The described method based on medium resolution daily land surface temperature and vegetation satellite data can successfully generate soil moisture and soil moisture index maps. Validation shows that the coefficient of determination between the satellite based maps and the ground measurements varies considerably. The scale difference between the satellite measurements and the *in situ* network can result in large variations in the soil moisture values depending on the location of the measurement station within the pixel. A larger satellite data set and more ground measuring stations that are better spread over the study area will give better calibration possibilities.

The aim of this study is to create an operational workflow to provide early warning for droughts or surface floods. For this purpose - the trend, instead of the exact value of soil moisture - is required. The error in the SM calculations is not dependent on the vegetation class, nor on the time of the year and therefor does not systematically influence the trend determination.

The soil moisture maps will be generated on a regular bases allowing for the spatially and temporarily continuous monitoring of soil moisture over a larger area. Soil moisture is a parameter that is used in the calculation of almost all drought indicators. Trends in the soil moisture, combined with other meteorological, climatological, geomorphological and pedological data are expected to provide information to help preventing drought and surface floods. The derived soil moisture maps can also be used as input it for irrigation and drainage schemes and land reclamation and improvement projects.

Figure 8: Soil moisture maps of 9 and 11 March 2014.



References

[1] B.W. Barett and G.P. Petropoulos. Satellite remote sensing of surface soil moisture. In G.P. Petropoulos, editor, *Remote Sensing of Energy Fluxes and Soil Moisture Content*, pages 85-120, CRC Press, Boca Raton, 2014.

[2] I. Hrnjak, T. Lukić, M. B. Gavrilov, S. B. Marković, M. Unkašević and I. Tošić. Aridity in Vojvodina, Serbia. *Theor Appl Climatol*, 115:323–332, 2014.

[3] A. Huete, C., Justice and W. van Leeuwen, editors. *MODIS vegetation index (MOD 13) algorithm theoretical basic document* 1-6, 1999

.

[4] B. van Leeuwen, Z. Tobak, Zs. Ladányi and V. Blanka. Satellite based soil moisture estimates for agricultural drought prediction. In M. Cvetković, K. Novak Zelenika and J. Geiger, editors, *Geomathematics – from theory to practice*, *6th HR-HU and 17th HU geomathematical congress*, pages 175-182, 2014.

[5] K. Mallick, B.K. Bhattacharya and N. K. Patel. Estimating volumetric surface moisture content for cropped soils using a soil wetness index based on surface temperature and NDVI. *Agricultural and Forest Meteorology*, 149(8): 1327 – 1342, 2009.

[6] N. R. Patel, R. Anapashsha, S. Kumar, S. K. Saha and V. K. Dadhwal. Assessing potential of MODIS derived temperature/vegetation condition index (TVDI) to infer soil moisture status. *Int. J. of Remote Sensing*, 30(1): 23 – 39. 2009.

[7] T. Schmugge. Measurements of surface soil moisture and temperature, In: R. J. Hobbs and H.A. Mooney, editors, *Remote Sensing of Biosphere Functioning*, pages 31-63, Springer-Verlag, New York, 1990.

[8] S. M. Vicente – Serrano, X. Pons-Fernandez and J. M. Cuadrat-Prats. Mapping soil moisture in the central Ebro river valley (northeast Spain) with Landsat and NOAA satellite imagery: a comparison with meteorological data. *Int. J. of Remote Sensing*, 25(20): 4325-4350. 2004.

[9] Z. Wang. *MODIS Land Surface Temperature Algorithm Theoretical Basis Document (LST ATBD)* 2-6.1999.